

國立臺灣大學電機資訊學院資訊工程學研究所

碩士論文

Department of Computer Science and Information Engineering

College of Electrical Engineering and Computer Science

National Taiwan University

Master Thesis

微網誌之短文情緒偵測：使用時間語境，社交，與回應資訊

Sentiment Detection of Micro-blogging Short Texts via Contextual, Social,
and Responsive Information



Chien-Yuan Wang

指導教授：林守德 博士

Advisor: Shou-De Lin, Ph.D.

中華民國 100 年 7 月

July, 2011

國立臺灣大學碩士學位論文
口試委員會審定書

微網誌之短文情緒偵測：使用時間語境，社交，與回應資訊

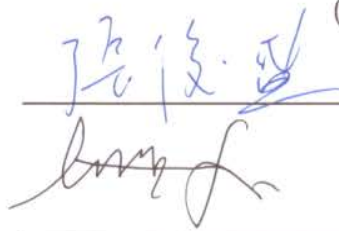
Sentiment Detection of Micro-blogging Short Texts via
Contextual, Social, and Responsive Information

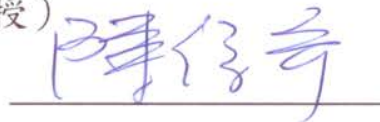
本論文係 王建元 君（學號R98922011）在國立臺灣大學資訊工程學系完成之碩士學位論文，於民國 100 年 7 月 15 日承下列考試委員審查通過及口試及格，特此證明

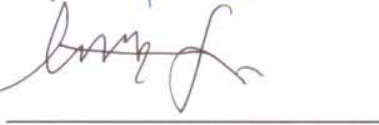
口試委員：

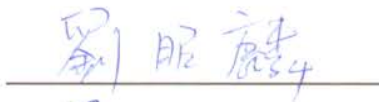


(指導教授)













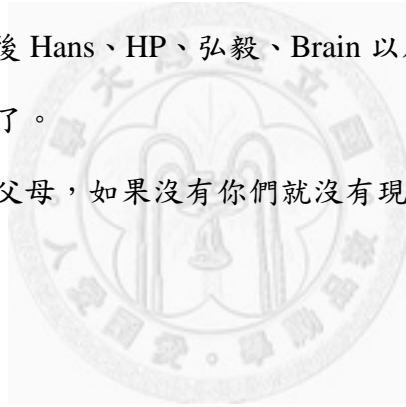
系主任

Acknowledgements

這篇論文的誕生要感謝很多人，首先要感謝林守德老師跟 Odd 學長在論文的研究上提供了很多建議與方向，特別謝謝 Odd 學長在百忙之中還抽空幫我看論文，還在口試 rehearsal 時提供很多寶貴的意見，還有恩勛，也非常謝謝你的幫忙與協助。此外感謝小 Ca 與 Poga 在我為研究困擾時，提供了很多寶貴的意見以及精神上的支持，當然還有亦辰、猴子、fish、西瓜，謝謝你們這兩年來的幫忙與陪伴。我們這一屆一起修課、出遊的時光歷歷在目。

昱志學長，謝謝你即使已經畢業還幫忙處理 server 和財產清點的問題。融璋，感謝你在我研究繁忙時能接手處理 server 以及採購的事，當然還有雅蓁，lab 的大小事務真的辛苦你了。最後 Hans、HP、弘毅、Brain 以及 lab 其他學長姐、學弟妹們，感謝這兩年來的關照了。

最後的最後，謝謝我的父母，如果沒有你們就沒有現在的我，謝謝你們一直以來的長久支持與關懷



摘要

微網誌是近來流行的社交平台，人們分享他們的生活或是關於某些事物的看法，這些資訊中蘊含著大量的情緒，也因此微網誌是一個好的資料來源，我們可以去分析這些人們的情緒感受，比如說像是平台上人們對某樣新產品的情緒，是正向或是負向。因此情緒偵測在微網誌上越來越有用，但因為微網誌平台的長度限制，每則發文最多只能有 140 個字，所以與其他類型的文章相較之下沒有足夠多的資訊可以使用來判斷。所以我們利用微網誌平台的特性，試著找出更多資訊能幫助微網誌上的短文情緒偵測。我們主要集中在三個方面 (a)時間語境, (b)社交, (c)回應訊息，並且提出三種方法可以利用這三方面的資訊，即Feature engineering based, Graphical model based, Markov-transition based。同時為了改善Memetube¹(原本的Pusic [1])情緒偵測的部分，它是一個能將微網誌的情緒音樂化的系統，主要基於六種基本情緒，所以我們集中在這六種基本情緒分類(anger, surprise, sadness, disgust, fear, joy) (Paul Ekman, 1992 [7])，它比起單純分成正向與負向情緒更具有挑戰性。

關鍵字: 情緒偵測，情緒分類

¹ <http://mslab.csie.ntu.edu.tw/memetube/>

Abstract

Micro-blog is a popular social platform recently, people shares their life, or comment about something, and all of this contain vast amount of sentiment, it's a good source we can use to analyze about the feeling of people, like what's the feeling of people about the new product, is positive or negative. Therefore, sentiment detection is more useful in micro-blog platform, but due to the length constraint, the maximum length of post in micro-blog is only 140 characters, there is not much information than other text genres. So we exploit the property of micro-blog platform to find more information to aid the sentiment detection of post in micro-blog. We focus on three aspects: (a) context, (b) social, (c) response, and propose three approaches, i.e., Feature engineering Based, Graphical model Based, and Markov-transition based , that can exploit the information from the three aspects. Meanwhile, for the purpose of improving the sentiment detection component of Memetube² system (original Pusic [1]), which is a platform that can musicalize the sentiment of micro-blogging messages for a given query, based on six basic emotion, so we focus on the six emotion (anger, surprise, sadness, disgust, fear, joy) (Paul Ekman, 1992 [7]), it's more challenging than positive and negative sentiment.

Keywords: sentiment detection, sentiment classification

² <http://mslab.csie.ntu.edu.tw/memetube/>

Table of Contents

| | |
|---|-----|
| Acknowledgements | i |
| 摘要 | ii |
| Abstract | iii |
| Table of Contents | iv |
| List of Figures..... | vi |
| List of Tables | vii |
| Chapter 1. Introduction..... | 1 |
| 1.1 Background..... | 1 |
| 1.2 Motivation and Purpose | 2 |
| 1.3 Research Statement | 3 |
| 1.4 Methodology Outline | 4 |
| 1.5 Contributions..... | 5 |
| 1.6 Paper Organization | 6 |
| Chapter 2. Related Works | 6 |
| Chapter 3. Methodology | 9 |
| 3.1 Original Supervised learning Model | 10 |
| 3.2 Feature engineering Based..... | 14 |
| 3.3 Graphical model Based..... | 16 |
| 3.4 Markov-transition Based | 19 |
| Chapter 4. Experiment | 23 |
| 4.1 Dataset preprocessing and Evaluation method | 23 |
| 4.2 Evaluation for original supervised learning model..... | 25 |
| 4.3 Evaluation for Feature engineering based approach | 26 |
| 4.4 Evaluation for Graphical model based approach | 28 |
| 4.5 Evaluation for Markov-transition based approach | 28 |

Chapter 5. Conclusion 32

Reference..... 33



List of Figures

| | |
|---|----|
| Figure 3-1 The relationship overview of methods mentioned in section 3.1~4 | 10 |
| Figure 3-2 The flowchart of generating base sentiment classifier using supervised machine learning..... | 10 |
| Figure 3-3 the overview of feature engineering based model, shows the features used in each instance for the three factors | 15 |
| Figure 3-4. the base CRF model, U: user {user_id}, S: sentiment of post {anger, surprise, sadness, disgust, fear, joy}, S_t connected with U_1 means the sentiment of U_1 's post at time t (P_t), w: word appeared in the post {word features, used in section 3.1 containing the n-gram, sentiment lexicon, PMI emotion words} | 17 |
| Figure 3-5. The base CRF model + context factor, consider the sentiment of previous post for each user..... | 17 |
| Figure 3-6. The base CRF Model + friend factor, if U_2 is a friend of U_1 , then the sentiment of friends' recent post may be correlated with the sentiment of users' post | 18 |
| Figure 3-7. The base CRF Model + response factor, the S_r represents the sentiment of response {anger, surprise, sadness, disgust, fear, joy}, correlated with the sentiment of post and words it contains. For example, the U_1 's post P_t has sentiment S_t , and it maybe has responses($R_1 \dots R_n$) containing sentiment $S_{r_1} \dots S_{r_n}$ | 18 |
| Figure 3-8. The base CRF model + Combine the three factors, context, friend, response. | 19 |
| Figure 3-9. The overview of Markov-transition based model..... | 19 |

List of Tables

| | |
|--|----|
| Table 4-1 Examples of emoticons belonging to the six basic sentiments in Plurk | 23 |
| Table 4-2 The result of base classifier for six class sentiment | 25 |
| Table 4-3 show the effect of features used in Liblinear L1R-LR..... | 26 |
| Table 4-4 the result of adding the emotion of three factors as feature (emoticon-labeled testing data, D1)..... | 27 |
| Table 4-5 the result of adding the prediction emotion of three factors as feature (emoticon-labeled testing data, D1)..... | 27 |
| Table 4-6 The result of CRF model using three factors. | 28 |
| Table 4-7 The results after adding addition info, evaluate using emoticon-labeled testing data (D1), Basic means the base classifier. | 29 |
| Table 4-8 the results of proposed three approaches in D2 data..... | 30 |
| Table 4-9 Binary classification for anger, the most improvement in accuracy is 0.7%, and in F-score is 1% | 31 |
| Table 4-10 Binary classification for surprise, the most improvement in accuracy is 4.3%, and in F-score is 6.7% | 31 |
| Table 4-11 Binary classification for sadness, the most improvement in accuracy is 0.2%, and in F-score is 0.4% | 31 |
| Table 4-12 Binary classification for disgust, the most improvement in accuracy is 1.9%, and in F-score is 2.4% | 31 |
| Table 4-13 Binary classification for fear, the most improvement in accuracy is 0.2%, and in F-score is 0.3% | 32 |
| Table 4-14 Binary classification for joy, the most improvement in accuracy is 1.9%, and in F-score is 1.9% | 32 |

Chapter 1. Introduction

1.1 Background

Micro-blogging is a newly emerging and popular social media in recent years, such as Twitter³, Plurk⁴, and Jaiku⁵, they allow users to share immediate but short messages with friends. People use more brief and colloquial expression to share things happened in the life, or discuss about any interesting things, e.g. fashionable product, political event, News happened today.

Generally, there is a lot of information posted in Micro-blog in each day. Some of them reveal personal status, and some of them are for social purpose. We can know the latest new of friend, and also can know the public news that happened in life and is discussed hotly by people. The people in Micro-blog platform ranges from general people to celebrity, like politician, famous singer, scientist, etc., we can observe different opinion coming from different type of people. Not only that, people comes from all over the world, we can know different opinions about the same issue from people of different country.

Besides, we can observe that there are explicit or implicit emotions contained in the posts. For instance, “*back from Amsterdam, was a really nice trip*” expresses explicit joy emotion in the post, and “*done with today's deadline*”, we can sense the relief and implicit joy, even without obvious emotion words in the posts, Therefore, if we further analyze for the sentiment involved in, we may know the public sentiment toward a product or the recent feeling of friends. Consequently, Micro-blog is a good source for

³ <http://twitter.com/>

⁴ <http://www.plurk.com/>

⁵ <http://www.jaiku.com/>

opinion mining and sentiment analysis.

Generally speaking, the sentiment detection is an issue of sentiment analysis. Sentiment detection is to detect subjective emotion involved in a given text. The text granularity can be document level, sentence level, or phrase level, and belong to different domain, e.g. blog, review, micro-blog, etc., they may have different writing style. Currently, there are two problems in sentiment detection, first, subjectivity classification, distinguishes whether a text is subjective opinion or objective fact, we can retrieve subjective opinions for further sentiment detection. Second, sentiment classification, classifies a given text to a sentiment from a predefined sentiment categories, for example, positive, neutral, negative. In this paper, we mainly focus on exploring the temporal context, social, and response information whether or not they can aid the sentiment classification of post in Micro-blog. Currently, the common research problem in sentiment detection of Micro-blog is to detect positive and negative emotion in a given post, and usually adopts machine learning method to solve this problem, and can achieve good performance.

1.2 Motivation and Purpose

Recently, a lot of sentiment analysis website for Micro-blog is appearing, for example, TwitterSentiment⁶[6], TweetFeel⁷. Given a query item, the service will analyze the sentiment about it in the Micro-blog platform. The sentiment detection component is important for such service, but due to the length-limited post of Micro-blog property, only 140 characters can be used for linguistic and textual analysis, besides, people can express emotion in an inconspicuous way, even human cannot

⁶ <http://twittersentiment.appspot.com/>

⁷ <http://www.tweetfeel.com/>

easily judge and categorize the emotion of post. Most of previous works ([4][5][6][8][9][10][11][22][23]) focus on the 140 characters length post, and try to find any useful linguistic feature or effective classification methods to solve this problem. So in addition to follow their research result, we try to find some related information that may be correlated to the post from property of Micro-blog platform, and propose approaches that use the found information in sentiment detection.

First, we observe the property of Micro-blog. Generally, the micro-blogging services possess some signature properties that differentiate them from conventional weblogs and forum. First, micro-blogging is *time-traceable*. The temporal information is crucial because contextual posts that appear close together are, to some extent, correlated. Second, the style of micro-blogging posts tends to be *conversation-based* with a sequence of responses. This phenomenon indicates that the posts and their responses are highly correlated in many aspects. Third, micro-blogging is *friendship-influenced*. Posts from a particular user can also be viewed by his/her friends and might have an impact on them (e.g. the empathy effect) implicitly or explicitly. Therefore, posts from friends in the same period may be correlated sentiment-wise as well as content-wise.

So in this paper, we focus on the three found information, i.e. the posts comes from temporal context, responses, and friendship, and propose three approaches that can exploit the found information, and verify these approaches whether or not they can aid the sentiment detection of the post. Besides, for diversity of human emotion, we focus on six basic sentiments {anger, surprise, sadness, disgust, fear, joy}, not just use positive and negative.

1.3 Research Statement

This research discusses how the response, context, and friendship information can

be exploited to improve the sentiment analysis given short text in Micro-blog posts.

Moreover, previous works ([4][6][8][9][10]) only focus on binary sentiment classification of positive and negative, and we focus on multi-class sentiment classification of the six basic sentiments.

1.4 Methodology Outline

Besides using the current approach for sentiment detection on Micro-blog, focus on feature engineering of the 140 characters and try different machine learning classifier model, we try to find some information that may be helpful. So from the three factors mentioned in the above, i.e. temporal context, social (friendship), responses, we can find some posts of the three factors, and they may be helpful for sentiment detection. Therefore, we propose three methods (1) Feature engineering based (2) Graphical model based (3) Markov-transition based, they can exploit the above three type of information for sentiment detection, and then we can verify that the three information are helpful if using appropriate approach.

At first, we follow current approach adopting supervised machine learning for sentiment detection in Micro-blog, so the Feature engineering based approach is the most intuitive way, we try to transform the information coming from the posts of the three factors to features, and see the features if helpful or not.

On the other hand, for modeling the sentiment correlation between target post and posts from three factors explicitly, the Graphical model based approach is the intuitive way for this aspect. We adopt probabilistic graphical model, it can uses graph to represent the relationship between variables, nodes in graph represent variables, and edges represent correlation between variables, so we can model the sentiment correlation between posts via graph explicitly. Besides, graphical model is ever used to

model the sentiment correlation between sentences in a blog article in previous works ([17]). Finally, the Markov-transition based approach is based on supervised machine learning like Feature engineering based approach, and we can choose any effective supervised learning method which can output probability distribution for given post. Besides, using the Markov-transition matrix of sentiment can consider the relationship between the three types of posts and target post, just like Graphical model based model can do. Further explanation, at first, we decide a base sentiment classifier, it's replaceable and need to output the sentiment probability distribution for the given post, and then we can use the base classifier to predict sentiment distribution of target post, meanwhile, we can also use sentiment transition matrix learned and posts from three factors to predict sentiment distribution of target post, finally, we merge the two sentiment distributions to output final adjusted sentiment distribution of target post.

1.5 Contributions

The detailed contributions in this paper are listed as follows:

- We study the problem of sentiment detection for six basic sentiments {anger, surprise, sadness, disgust, fear, joy} in Micro-blog, It's natural to study more detailed sentiment further after positive and negative sentiments are explored widely.
- We propose three approaches to exploit temporal context, friendship, and response to improve sentiment classification, i.e. (1) Feature engineering based (2) Graphical model based (3) Markov-transition based. The Markov-transition based approach has the most improvement after exploiting the three type of information.
- The Markov-transition based approach can be applied to the sentiment detection

component of Memetube⁸ system, make it more accurate than original keyword matching approach.

1.6 Paper Organization

This paper is organized as follows. In Section 2: Overview the related works about sentiment detection in Micro-blog platform. In Section 3: introduction the method we proposed that use the temporal context, friendship, response. In Section 4: experiment our model on Plurk data. In Section 5: Conclusion and future work.

Chapter 2. Related Works

In recent years, the related works about sentiment analysis of Micro-blog domain is increasing gradually because of the role of Micro-blog, it is not only a web 2.0 website, which facilitates the information sharing, but also a social media service, has more communications than blog between people around the world. Therefore, Micro-blog is a good research source, which containing a lot of continuously growing data, and can be used for analysis for public sentiment.

The common sentiment detection problem of Micro-blog in most related works is that given a length-limited post of Micro-blog, could the system tell which sentiment the post belongs to? The current approach (Go et al. 2009[6], Li et al. 2009[23], Barbosa and Feng 2010[8], Bermingham and Smeaton 2010[9], Bifet et al. 2010[10], Davidov et al. 2010[11], Pak and Paroubek 2010[22], Sun et al. 2010[4]) for sentiment detection of Micro-blog focus on two aspects: first, feature engineering for a variety of textual and linguistics features, like n-gram, POS tagging, prior subjectivity and polarity

⁸ <http://mslab.csie.ntu.edu.tw/memetube/>

of word, punctuation, word pattern, or generic feature, e.g. URL, hashtag...etc.; second, try different classification techniques, usually belongs to two types: unsupervised machine learning algorithm, e.g. keyword matching; and supervised machine learning algorithm, e.g. Naive Bayes, Maximum Entropy, SVM.

Generally speaking, supervised machine learning needs enough training data to get a better model. Besides, if there is more training data, the high coverage of potential words used in posts can generate more robust model for use in applications. The current approach for labeling the vast amount of training data for use is using emoticons, it's a facial expression composed of characters, e.g. :-), :-o. Go et al. (2009)[6] use emoticons to label twitter data as positive or negative sentiment automatically, the method introduced by Read (2005)[12], if a post contains emoticon, it can be treated as like the author of the post label the emotion of the post by himself using emoticon, therefore, Go et al. can get abundant of training data for supervised machine learning algorithm. Recently, some researchers also follow this labeling approach (Chen et al. 2010 [5]; Davidov et al. 2010[11]; Sun et al. 2010 [4]; Bifet and Frank 2010 [10]; Pak and Paroubek 2010[22]). In our research, we follow this approach to get enough training data, too.

So far, The above related works only consider the information contained in the post, recently, some works start to try improving sentiment detection from different angle, they focus on topic-dependent sentiment detection and take other related information into consideration. Jiang et al. 2011[13], focus on sentiment detection of posts about a specific target in Micro-blog, e.g. Obama, Google, Ipad, adopting positive and negative sentiment as label, they propose multiple rules to generate target-dependent features, for example, "*I love ipad*", then take related word patterns ("*love_ipad*") as features, besides, also consider related posts into consideration, like

retweets, responses, posts about the same target of the same user. Calais et al. 2011[14] adopt transfer learning to solve the sentiment detection about a specific topic, they first learn the bias of users toward the specific topic, and transfer the information learned to the problem of sentiment classification of positive and negative sentiment in Micro-blog. So from the above recent works mentioned, we can know that adopting relevant information about the post may be a consistent direction for sentiment detection in Micro-blog. Currently, most of related works focus on positive and negative sentiment, their accuracy ranges from 60%~90% depending on different setups of data and classifier. Because of the diversity of human emotion, exploring more sentiment types should be helpful for emotion-related analysis and application in the future. Bollen et al. (2010)[15] analyze the relation between public sentiment of twitter platform and social, economic events in real word with six emotion: Tension, Depression, Anger, Vigour, Fatigue, and Confusion, and found the events do have an significant effect for the six emotion dimension in twitter. So we believe that adopting more detailed sentiment to sentiment detection of Micro-blog will be useful for further sentiment analysis.

To summarize, the current research problem of sentiment detection in Micro-blog has two forms, topic-dependent or topic-independent, and they usually consider positive and negative sentiments. Previous works of topic-independent form mainly focus on the information contained in the length-limited post, so we try to find related information about the post into consideration, and then apply the three kinds of additional information found (i.e., contextual, response and friendship information) for sentiment recognition. Just like recent works in topic-dependent form, they also try to find related information into consideration, and we are encouraged for such approach. Besides, we adopt six basic emotions: Anger, Surprise, Sadness, Disgust, Fear, Joy, it's more challenging for consideration of the further detailed sentiment of human.

Chapter 3. Methodology

The major challenge in the micro-blog sentiment detection task is that the length of each post is limited (i.e., posts on Plurk are limited to 140 characters), besides, people can convey the sentiment in a subtle way without obvious emotional words, e.g. “someone keeps texting me and I don’t know the person..”, we can be aware that an implicit anger and annoyed emotion is involved in the sentence. Consequently, there might not be enough information for a sentiment detection system to exploit. To solve this problem, we propose to utilize the three types of information mentioned earlier, i.e., temporal context, friendship, response. We will give a detailed explanation in section 3.1.

And then, we will show our proposed three approaches: (1) Feature engineering Based (section 3.2) (2) Graphical model Based (section 3.3) (3) Markov-transition Based (section 3.4), for using the context, friendship, response information to improve the sentiment detection of Micro-blog. The overview of the methods mentioned in following section shows in the following Figure.

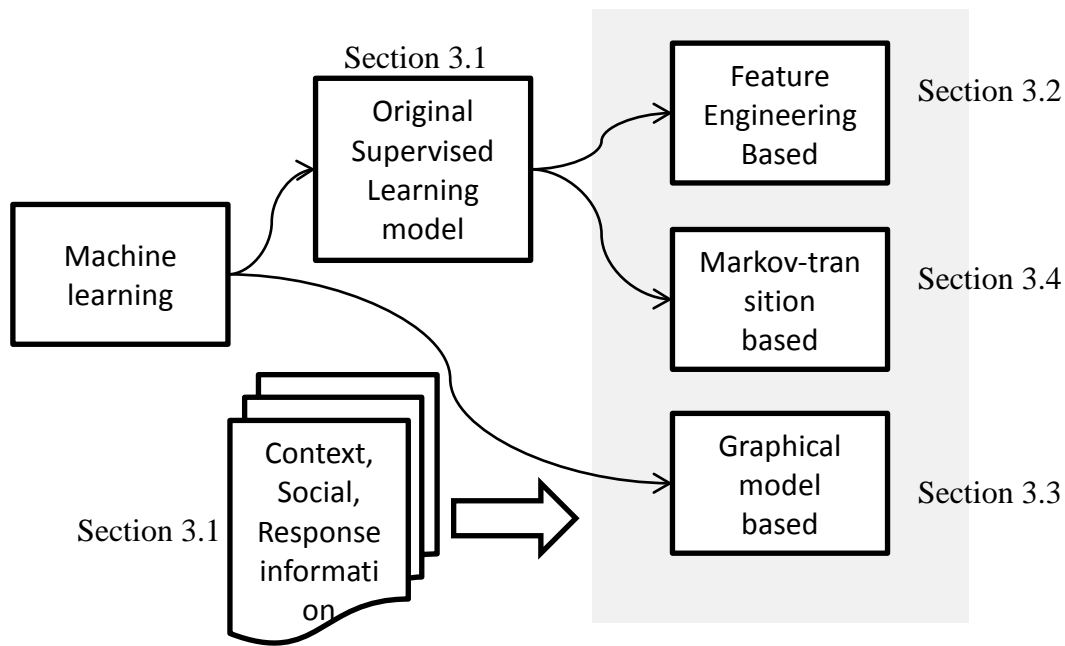


Figure 3-1 The relationship overview of methods mentioned in section 3.1~4

3.1 Original Supervised learning Model

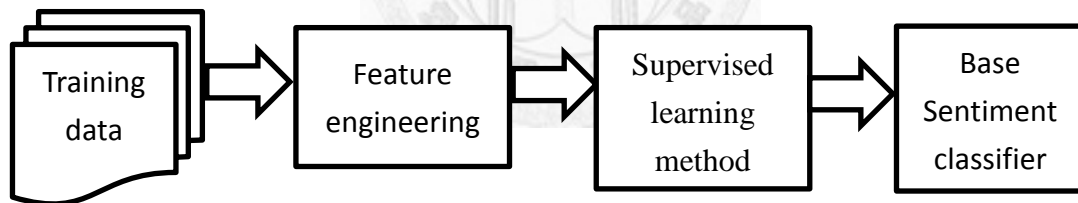


Figure 3-2 The flowchart of generating base sentiment classifier using supervised machine learning

The most basic method for sentiment classification adopting supervised machine learning technique is shown in Figure 3-2. We will use the base sentiment classifier in Feature engineering based and Markov-transition based approaches. Just like Figure 3-2 Shows, from training data, at first, we need to do feature engineering, propose features that are effective for sentiment classification, the final found effective features are

called *base feature* in the following sections, and then we decide appropriate supervised machine learning method to train a base sentiment classifier out.

The features used in classification model are listed in below:

- N-gram based feature

The most common used feature in sentiment detection, it usually can perform a nice result in related works ([6][8][9][10][11][22]). So we use the n-gram of the sentence in training data as binary feature. In here, we use unigram and bigram, and the words appearing less than defined frequency in the data are filtered.

- Emotion dictionary feature

The emotion words are relevant to the sentiment of the sentence intuitively, so we take the emotion word as feature. There are some famous emotion dictionaries, for Chinese, e.g. NTUSD [2], and for English, e.g. AFINN [3]. In addition to direct affective word, like joy, anger,...,etc., there are some words, they are relevant to arouse the emotion of human, called indirect affective word, like earthquake, tsunami, so we collect natural disaster words from Wikipedia to expand the emotion dictionary. In here, we use the word in the dictionary as binary feature.

- Emotion word extracted by variance of PMI

Use the point-wise mutual information, we can extract the words, they are relevant to emotion, in here we use the variance of point-wise mutual information introduced by ([4] [17]) to retrieve top N word of each emotion back as binary feature. The definition of variance of PMI is as follows:

$$Co(s, w) = C(s, w) \times \log_e \frac{P(s, w)}{P(s)P(w)}$$

where $C(s, w)$ means the count of sentiment and word appearing concurrently, and $P(s)$ means the probability of sentiment appearing in the data, $P(w)$ means the probability of word, finally, $P(s, w)$ means the joint probability of sentiment and

word, and normalize to 0~1, get $Co'(s,w)$.

$$Co'(s,w) = \frac{Co(s,w) - Co_{min}}{Co_{max} - Co_{min}}$$

Because we only care about part of speech, which are relevant to emotion, so we segment the sentence, and use POS tagger [16] to keep only adjective, noun, verb words to the extraction process using PMI.

- Generic feature

In addition to the above feature, we use the generic feature like the length of sentence, normalize to 0~1, by divide by 140, the maximum sentence length allowed in Micro-blog system. And the question mark(?) and exclamation mark(!) in punctuation are relevant to emotion, we use the count of times appeared in the sentence as feature. Besides, the post may be a share post, containing a HTTP URL address, we use meta tag URL to replace the pattern, and use it as a binary feature.

Use the above feature, we can train a base sentiment classification model adopting any appropriate supervised machine learning model, e.g. NaiveBayes, SVM, and then it allow us to produce a probability distribution of sentiments for a given post p , denoted as S_p . We will try to find the most effective machine learning model, and then apply it as base sentiment classifier for use in Feature engineering based and Markov-transition based approaches.

Before introducing the three approaches proposed for the three type of information, i.e. temporal context, friendship, response. , we first explain the three factors in detail. The temporal context factor, it is assumed that the sentiment of a Micro-blog post is correlated with the sentiments of the author's previous posts (i.e., the 'context' of the post); and the friendship factor, we assume that the friends' emotions are correlated with each other. This is because friends affect each other, and they are more likely to be in the same circumstances, and thus enjoy/suffer similarly. Our hypothesis is that the

sentiment of a post and the sentiments of the author's friends' recent posts might be correlated; finally, the response factor, we believe the sentiment of a post is highly correlated with (but not necessary similar to) that of responses to the post. For example, an angry post usually triggers angry responses, but a sad post usually solicits supportive responses.

Besides, for friendship factor, how to choose friends whose sentiment of recent posts may be correlated from higher probability to low probability? Generally, friends in the same circumstances should experience the same thing with higher probability, so the posts in the same period may be correlated, and the sentiment contained may be correlated, too. For example, students may publish posts containing unhappy emotion in the interval of final exam, and after final exam is over, they may publish happy posts, so if user and friends are classmates, they are in the same circumstances, they will have similar emotion patterns, in here, we can treat sentiment transition as an emotion pattern, in the example, there is a transition from unhappy to happy. Therefore, we assume that if the emotion pattern is more similar, the sentiment of the friends' recent posts is more correlated. We regard one's sentiment transition matrix M_c as emotion pattern of each user, where the $(i, j)^{th}$ element in M_c is the conditional probability from the sentiment of the previous post to that of the current post, and then we propose to give priority to friends with similar emotional patterns.

To achieve our goal, we first learn every user's contextual sentiment transition matrix M_c from the data. In M_c , each row represents a distribution that sums to one; therefore, we can compare two matrixes M_{c_1} and M_{c_2} by averaging the symmetric KL-divergence of each row. That is,

$$Distance(M_1, M_2) = Average_{i=1}^n KL(Row(M_1, i), Row(M_2, i)).$$

Two persons are considered as having similar emotion pattern if their contextual

sentiment transition matrixes are similar. That means the distance is small.

After explanation for the three factors, we continue to talk about approaches using the three factors in the following section.

3.2 Feature engineering Based

If consider taking the three additional information into original base sentiment recognition mechanism of supervised learning, the most intuitive way is to add the features coming from the three factors, and then see the features if helpful or not. We take the emotion of post from the three factors as features. Note that if take them as feature, the emotions of posts from three factors must be known in advance. So in the experiment, we assume that we already know the emotion of three factors, otherwise, if using the prediction emotion of the three factors as feature, it will be dependent on the accuracy of prediction result, for example, if the accuracy of prediction emotion is extremely low, that means it may bring in fault emotion as features. In here, we can view the experiment as verification for the three factors, if we know the emotion of previous post, or the emotion of friends' previous post, or the emotion of responses, and then add them as feature, will they improve the base sentiment recognition? Just like the overview in the below:

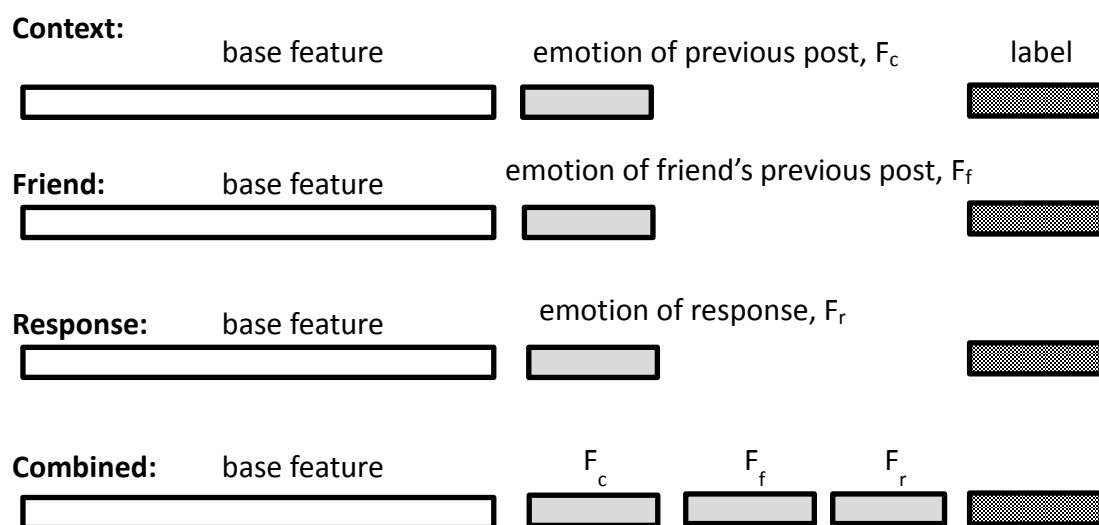


Figure 3-3 the overview of feature engineering based model, shows the features used in each instance for the three factors

The base feature in Figure 3-3 is the features used in section 3.1, contains n-gram, sentiment lexicon, PMI emotion words, and generic features. The emotion feature of previous post F_c represents the appeared times of each emotion in previous posts within the specific time interval, $[N_{\text{anger}}, N_{\text{surprise}}, N_{\text{sadness}}, N_{\text{disgust}}, N_{\text{fear}}, N_{\text{joy}}]$, for example, if the post contains a previous post containing anger emotion within specified hour, and the emotion feature is $[1, 0, 0, 0, 0, 0]$. So for temporal context factor, we can decide (1) amount of previous posts, (2) the specific time interval to generate the emotion feature; and for friendship factor, we can decide (1) top N friend, they are selected by methods mentioned in section 3.1, (2) amount of friends' previous posts (3) the specified time interval; And finally, the response factor, we can decide (1) first N responses of the post. So use the parameters defined for each factor, we can generate emotion features for each post.

After generating features for each factor as Figure 3-3, we adopt the most effective supervised machine learning method in section 3.1 to re-train the sentiment classifier, and see if adding the features came from the three factors will be helpful or not. We can

the compare the result with sentiment classifier with only base feature.

3.3 Graphical model Based

For modeling the hypothetical sentiment correlation between target post and the posts of three factors, which is mentioned in section 3.1, we can use probabilistic graph model to deal with it.

We choose to use CRF (condition random field) because it is ever used as context-level classifier to model the sentiment correlation for a sequence of sentences in the article for blog domain [17],so similarly, we treat the response of post as context, and the same to previous post, and friends' recent post The idea of CRF, is to model the $P(Y|X)$ in an proposed undirected graph, where the nodes in graph belongs to two disjoint set X and Y , where X is the observed variables, Y is the target variables, all nodes in graph $\in X \cup Y$.

The general parameter learning of CRF is

$$P(Y|X, \theta) = \frac{1}{Z(X, \theta)} \exp\left[\sum_i \theta_i F_i(Y, X) \right]$$

Where X is observed variables, Y is target variables, θ is the parameters to be learned from training data, $Z(X, \theta)$ is normalization factor, $F_i(Y, X)$ is the feature function.

The inference for target variables is given X, θ , solve

$$Y' = \operatorname{argmax}_y P(Y|X, \theta)$$

We use approximate inference adopting Gibbs sampling to solve it.

The proposed graph for modeling the three factors shows in below Figure 3-5~Figure 3-8, Figure 3-4 is the base graph which not yet models the three factors. Note that the target variable is the sentiment of post, and others are treated as observed variables.

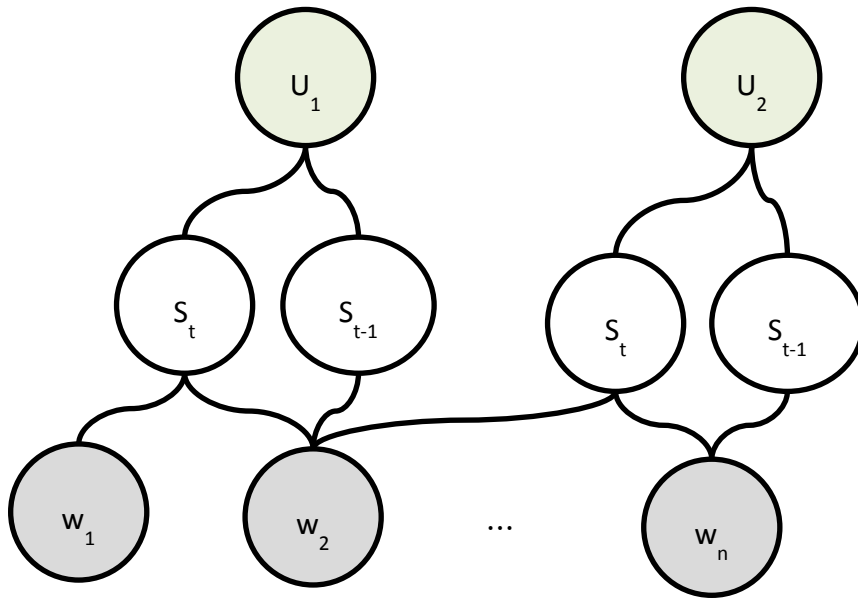


Figure 3-4. the base CRF model, U: user {user_id}, S: sentiment of post {anger, surprise, sadness, disgust, fear, joy}, S_t connected with U_1 means the sentiment of U_1 's post at time t (P_t), w: word appeared in the post {word features, used in section 3.1 containing the n-gram, sentiment lexicon, PMI emotion words}

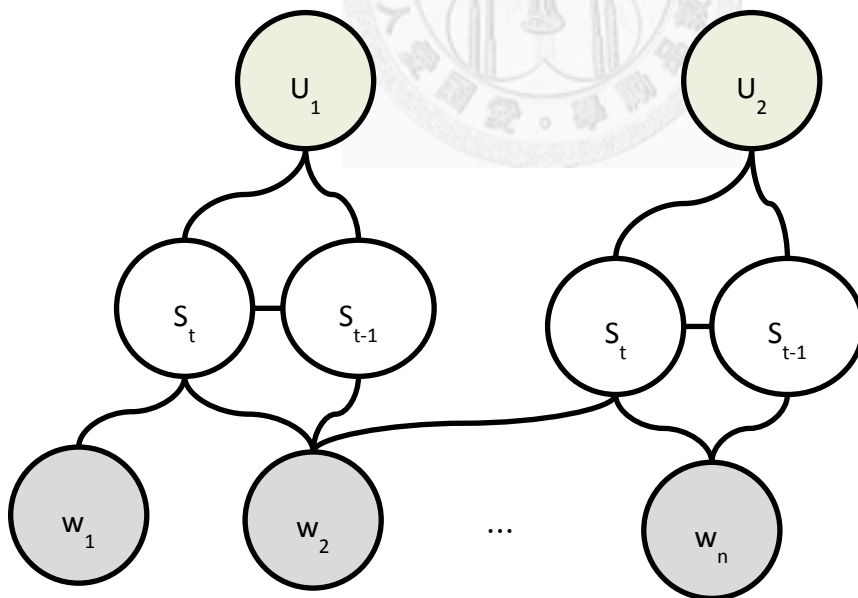


Figure 3-5. The base CRF model + context factor, consider the sentiment of previous post for each user

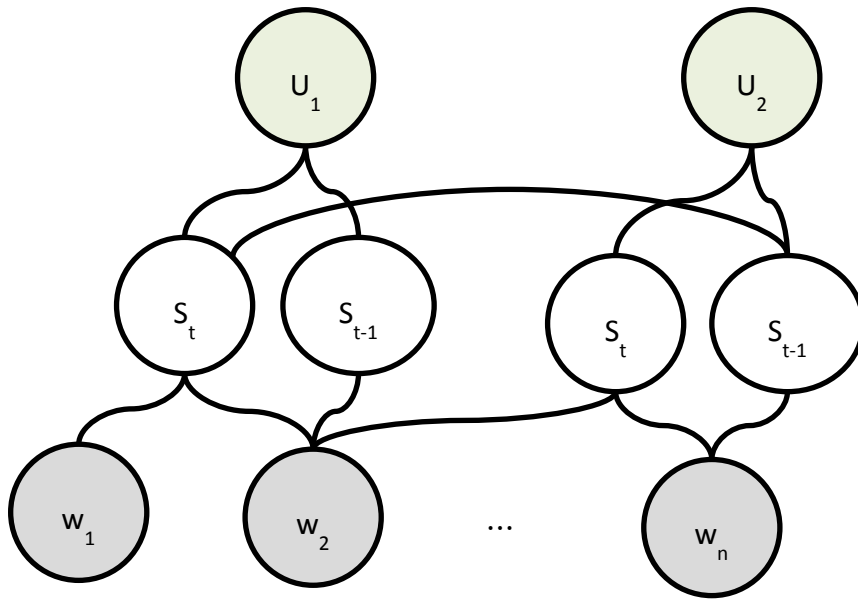


Figure 3-6. The base CRF Model + friend factor, if U_2 is a friend of U_1 , then the sentiment of friends' recent post may be correlated with the sentiment of users' post

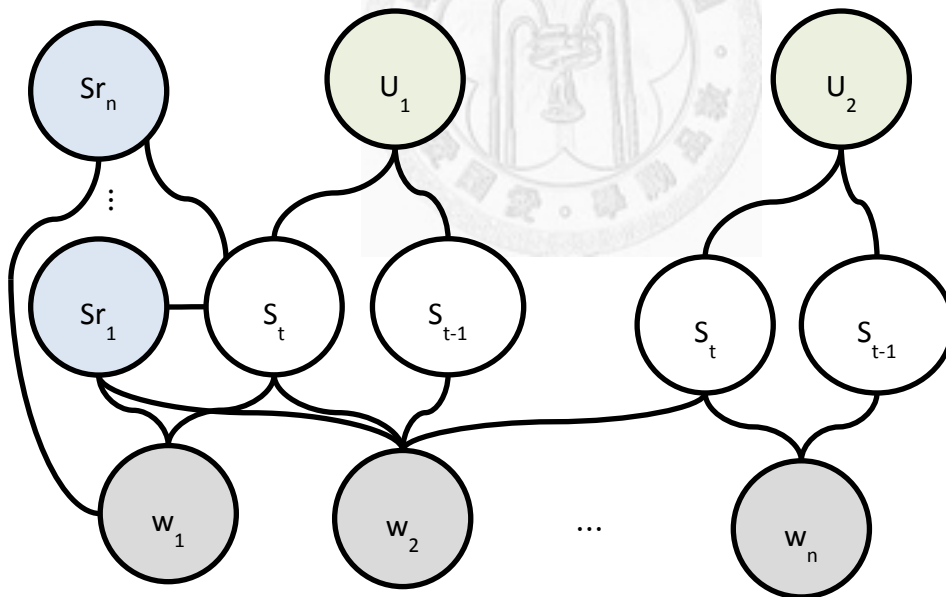


Figure 3-7. The base CRF Model + response factor, the S_r represents the sentiment of response {anger, surprise, sadness, disgust, fear, joy}, correlated with the sentiment of post and words it contains. For example, the U_1 's post P_t has sentiment S_t , and it maybe has responses($R_1 \dots R_n$) containing sentiment $S_{r_1} \dots S_{r_n}$.

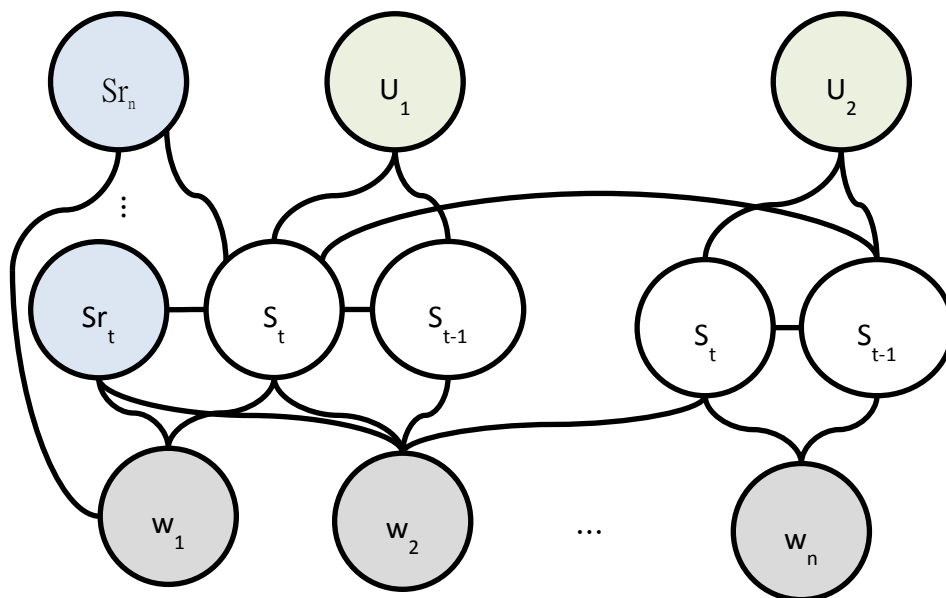


Figure 3-8. The base CRF model + Combine the three factors, context, friend, response.

3.4 Markov-transition Based

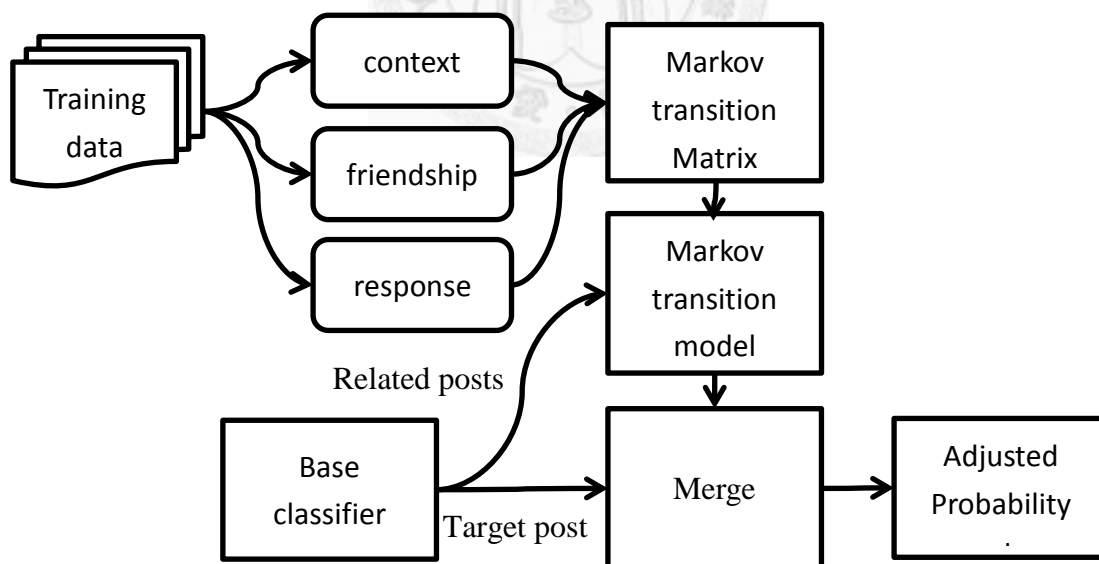


Figure 3-9. The overview of Markov-transition based model

Just like the overview (Figure 3-9) shows, first, from the training data, we can learn sentiment Markov-transition matrix for the three factors, i.e. context, friendship,

response, for example, the joy response often reply to the joy post than other emotion post, so we can use the conditional probability $P(\textit{Sentiment}_{post}|\textit{Sentiment}_{response})$ to model the relationship of the tendency, therefore, we can construct a sentiment transition matrix. And the base classifier in Figure 3-9 is the sentiment classifier trained in section 3.1, given a post or a response, it can output a sentiment probability distribution S_p . And then we can merge the S_p of target post with the result of S_p of responses (related posts) multiplied by the sentiment transition matrix, and get an adjusted probability distribution.

In the experiment, we can compare the result of such Markov-transition based approach with original sentiment classifier, to see whether or not the three factors are helpful.

We discuss how to exploit the three type of information in detail below.

Response Factor

We propose to learn the correlation patterns of sentiments from the data and use them to improve the recognition. To achieve such goal, from the data, we learn the probability $P(\textit{Sentiment}_{post}|\textit{Sentiment}_{response})$, which represents the conditional probability of a post given responses. Then we use such probability to construct a transition matrix M_r , where $M_{r_{ij}}=P(\textit{Sentiment}_{post} = j | \textit{Sentiment}_{response} = i)$.

With M_r , we can generate the adjusted sentiment distribution of the post S'_p as:

$$S'_p = \alpha \times \frac{\sum_{i=1}^k W_{r_i} S_{r_i} M_r}{k} + (1 - \alpha) S_p$$

where S_p denotes the original sentiment distribution of the post, and S_{r_i} is the sentiment distribution of the i^{th} response determined by the abovementioned base sentiment-detection model approach. In addition, $W_{r_i} = 1/(t_{response_i} - t_{post})$ represents the weight of the response since it is preferable to assign higher weights to closer responses. There is also a global parameter α that determines how much the

system should trust the information derived from the responses to the post. If there is no response to a post, we simply assign $S'_p = S_p$.

Context Factor

. We assume that, for each person, there is a sentiment transition matrix M_c that represents how his/her sentiments change over time. The $(i, j)^{th}$ element in M_c represents the conditional probability from the sentiment of the previous post to that of the current post: $P(\text{Sentiment}(P_t) = j \mid \text{Sentiment}(P_{t-1}) = i)$.

The diagonal elements stand for the consistency of the emotion state of a person. Conceivably, a capricious person's diagnostic M_{cii} values will be lower than those of a calm person. The matrix M_c can be learned directly from the annotated data.

Let S_t represent the detected sentiment distribution of an existing post at time t . We want to adjust S_t based on the previous posts from $t - \Delta t$ to t , where Δt is a given temporal threshold. The system first extracts a set of posts from the same author posted from time $t - \Delta t$ to t and determines their sentiment distributions $\{S_{t_1}, S_{t_2}, \dots, S_{t_k}\}$, where $t - \Delta t < t_1, t_2, \dots, t_k < t$ using the same classifier. Then, the system utilizes the following update equation to obtain an adjusted sentiment distribution S'_t :

$$S'_t = \alpha \times \frac{\sum_{i=1}^k W_{t_i} S_{t_i} M_c}{k} + (1 - \alpha) S_t$$

where $W_{t_i} = 1/(t - t_i)$. The parameters W_{t_i} , k , α are defined similar to the previous case. If there is no post in the defined interval, the system will leave S_t unchanged.

Friendship Factor

We can treat the friends' recent posts in the same way as the recent posts of the author, and learn the transition matrix M_f , where $M_{f_{ij}} = P(\text{Sentiment}_{user}(P_t) = j \mid \text{Sentiment}_{user's\ friend}(P_{t-1}) = i)$, and apply the technique proposed to improve the recognition accuracy.

$$S'_t = \alpha \times \frac{\sum_{fr=1}^F W_{fr} \times \left(\frac{\sum_{i=1}^k W_{t_i} S_{t_i} M_f}{k} \right)}{F} + (1 - \alpha) S_t$$

Where $\left(\frac{\sum_{i=1}^k W_{t_i} S_{t_i} M_f}{k} \right)$ can be regarded as the prediction sentiment distribution using friends' recent posts, W_{fr} is the weight of the friend, F means the amount of friends used.

After a set of similar friends are identified, which is mentioned in section 3.1, their recent posts (i.e., from $t - \Delta t$ to t) are treated in the same way as the posts by the author, and the weight of friend W_{fr} is the reciprocal of $Distance(M1, M2) + 1$, the more similar friend has higher weight, we use the method proposed previously to fine-tune the recognition outcomes.

Combined: Response Factor + Context Factor + Friendship Factor

Use the above three information mentioned, we can combine the above equations as

$$S'_p = \alpha_c \times \frac{\sum_{i=1}^k W_{t_i} S_{t_i} M_c}{k} + \alpha_f \times \frac{\sum_{fr=1}^F W_{fr} \times \left(\frac{\sum_{i=1}^k W_{t_i} S_{t_i} M_f}{k} \right)}{F} + \alpha_r \times \frac{\sum_{i=1}^k W_{r_i} S_{r_i} M_r}{k} + (1 - \alpha_c - \alpha_f - \alpha_r) S_p$$

This final adjusted sentiment distribution use the information of original post, and the information come from temporal context, friendship, and response.

To summarize, The Markov-transition based approach possess some properties list in blow:

- It adopts the supervised learning model as base model like Feature engineering based approach, so it has the advantage that we can replace any effective supervised learning model which can output prediction probability.
- It can specify a time interval, which can bring in more related posts from the three factors, and also can specify other control parameters, e.g. amount of previous post,

top N friend, and first N response.

- The Markov-transition matrix can model the sentiment correlation between the posts from three factors and the target post, just like Graphical model based approach can do. .

Chapter 4. Experiment

4.1 Dataset preprocessing and Evaluation method

We collect the posts and responses from every effective Chinese user, users with more than 10 messages, of Plurk from January 31st to May 23rd, 2009. We use emoticons (Table 4-1) to extract posts and the sentiment of each post is labeled automatically using the emoticons. This is similar to what many people have proposed for evaluation mentioned in related works. (Note that the emotions are removed from posts after labeling).












| Sentiment | emoticon | code | emoticon | code | emoticon | code |
|-----------|---|-----------|---|--------------|---|-------|
| anger |  | X-(|  | (angry) | | |
| surprise |  | :-o | | | | |
| sandess |  | :'-(|  | (tears) | | |
| disgust |  | (annoyed) | | | | |
| fear |  | :-(or :(| | | | |
| joy |  | :-) or :) |  | :-D or :D |  | (LOL) |
| |  | :~)) | | | | |

Table 4-1 Examples of emoticons belonging to the six basic sentiments in Plurk

We use data from January 31st to April 30th as training set, May 1st to 23rd as testing data. And then, there two kind of pre-processing approaches:

(1) It's a common approach, in order to eliminate sentiment bias for predicting the sentiment of a given post, we retrieve same amount of posts for each sentiment for use in the training and testing data. For the purpose of observing the result of using the three factors, we filter the users without friends, the posts without responses, and the posts without previous post in 24 hour in testing data. Finally the amount of training and testing data is (53514, 2598) posts. This data we called *D1* in the following section.

(2) Retrieve data by user, not by same amount of posts for each sentiment. Because the graphical model based approach we proposed needs the user information just like the proposed graph shows in Figure 3-5~Figure 3-8, so if we retrieve training and testing data by user, it can learn the unique sentiment correlation between posts for each user, for example, the unique sentiment transition from previous post to current post of each user. And for make sure that every post has the three factors information as far as possible in training or testing data, we choose the user has training data, in which $\geq 50\%$ containing responses, and testing data, in which $\geq 50\%$ containing responses, too. Finally, get 982 users, 14751 posts in training data, 6557 posts in testing data, 71.3% users have friends in this data, and because retrieving data by user, so testing data surely contains previous posts. The distribution of testing data is that anger: 13.0%, surprise: 5.9%, sadness: 18.9%, disgust: 7.1%, fear: 9.7%, joy: 45.3%. This data we called *D2* in the following section.

And about the evaluation methods, we use the accuracy, F-score, or RMSE to evaluate our models, the F-score represents the average F-score of each sentiment.

Note that our research statement is “how the response, context, and friendship information can be exploited to improve the sentiment analysis”, so we mainly observe

that in the given data $D1$ or $D2$, whether or not our proposed three approaches can exploit three kinds of information to improve the result of sentiment classification, that represents that the result using the three kinds of information is better than the one without using.

4.2 Evaluation for original supervised learning model

First, we compare the result of different base sentiment classifier, and then we can find the most effective base sentiment classifier to continue doing the following experiment of using three information, context, friendship, response in Section 4.3 and 4.5. The result of base classifier shows in below:

| | Keyword Matching | LM | NaiveBayes | Libsvm C-SVM(linear) | Liblinear L2R-LR | Liblinear L1R-LR |
|----------|------------------|-------|------------|----------------------|------------------|------------------|
| Accuracy | 0.19 | 0.343 | 0.301 | 0.375 | 0.385 | 0.393 |
| F-score | 0.189 | 0.343 | 0.302 | 0.369 | 0.380 | 0.386 |

Table 4-2 The result of base classifier for six class sentiment

In Table 4-2, the keyword matching method is the original emotion detection method of Memetube, use six-class affective keywords from Wordnet-affect sentiment lexicon [18], and for evaluation, we use the majority vote to decide the emotion of post, if there is a tie or no keywords matched, we guess an emotion from candidates randomly. The language model classifier (LM) utilizes bigram as feature, and we use the computational linguistics tool kit, LingPipe [19] to implement it. Finally, the rest of classifiers, i.e. NaiveBayes, LibSVM, Liblinear ([20][21]) use the features mentioned in section 3.1(i.e. unigram, bigram, emotion dictionary, PMI words, generic features) to train a sentiment classifier.

We can see that the best base classifier is Liblinear L1R-LR, note that the main

reason the accuracy is not extremely high is that we are dealing with 6 classes. When we train a classifier for positive/negative, the classifier L1R-LR reaches 80.7% in accuracy, 80.8% in F-score, which is competitive to the state-of-the-art algorithms as shown in the related work section.

The features used in Liblinear L1R-LR shows in Table 4-3, we can see that actually unigram+bigram can perform nice result even without other features.

| Feature | accuracy |
|---|-----------------|
| Unigram+bigram | 0.389 |
| Unigram+bigram+emotion_dictionary | 0.393 |
| Unigram+bigram+emotion_dictionary +PMI emotion words | 0.392 |
| Unigram+bigram+emotion_dictionary+ generic_feature | 0.392 |
| Unigram+bigram+emotion_dictionary +PMI emotion words + generic_feature | 0.393 |

Table 4-3 show the effect of features used in Liblinear L1R-LR

4.3 Evaluation for Feature engineering based approach

The results of adding additional information of the three factors as features are shown in below (Table 4-4). The Basic in table, means only use the base feature. Note that the emotions of posts from three factors are known in the experiment.

The experiment shows that if we can know the emotion of previous post, friends' recent post, and responses, we can get an improvement over Basic classifier, and the Combined result has the most improvement.

| Liblinear L1R-LR | Basic | Context | Friend | Response | Combined |
|---------------------|-------|---------|--------|----------|----------|
| Accuracy | 0.393 | 0.420 | 0.397 | 0.445 | 0.470 |
| F-score | 0.386 | 0.416 | 0.392 | 0.438 | 0.466 |

Table 4-4 the result of adding the emotion of three factors as feature (emoticon-labeled testing data, D1)

If we try to use the prediction emotion of the posts from the three factors as features, the result shows in below:

| Liblinear L1R-LR | Basic | Context | Friend | Response | Combined |
|---------------------|-------|---------|--------|----------|----------|
| Accuracy | 0.393 | 0.387 | 0.397 | 0.393 | 0.395 |
| F-score | 0.386 | 0.383 | 0.391 | 0.386 | 0.390 |

Table 4-5 the result of adding the prediction emotion of three factors as feature (emoticon-labeled testing data, D1)

The prediction emotion features are generated by the Basic sentiment classifier, so the correctness of features of the three factors depends on the accuracy of Basic sentiment classifier, 0.393. Consequently, because the accuracy is not high enough, just as the Table shows, it causes an unstable result, retrograde in Context factor, and improve in Friend and Combined factor.

So although the Feature engineering based approach can get an improvement for all three factors if the emotions of posts from the three factors are known, it cannot get an ideal result if using the predicted emotions. So the Feature engineering based approach can exploit the information of the three factors to improve sentiment classification if the predicted emotion features of the three factors are correct enough, that depends on the

accuracy of basic sentiment classifier.

4.4 Evaluation for Graphical model based approach

Given D2 data, the result of general CRF model using the three factors shows in below. The response factor gets the most improvement in both evaluation methods. Other factors only get little improvement in accuracy, and bad result in RMSE.

| CRF | Basic | Context | Friend | Response | Combined |
|----------|-------|---------|--------|----------|----------|
| Accuracy | 0.508 | 0.512 | 0.511 | 0.518 | 0.515 |
| RMSE | 0.662 | 0.663 | 0.662 | 0.657 | 0.667 |

Table 4-6 The result of CRF model using three factors.

We infer that the previous posts, friends' recent posts are not like responses which have explicit correlation with target post, the target post and responses are in the same topic, and the previous posts or friends' recent posts are maybe not. So it cannot perform well or even get a bad result if we learn and fit the sentiment correlation using graphical model approach.

So the graphical model based approach can exploit the response factor to improve the sentiment classification, and other factors cannot perform well in such way.

4.5 Evaluation for Markov-transition based approach

Given D1 data, the result of Markov-transition based approach is shown in Table 4-7. Note that the sentiment distributions of the posts from three factors are using the predicted sentiment distribution generated from basic sentiment classifier.

The results show that considering all three additional factors can achieve the best results. The results show decent improvement over the base classifier, around 4.8%

improvement in accuracy, 5.1% in F-score, and when achieving the accuracy 0.441, the weight of Combined is $\alpha_c:0.2$, $\alpha_f:0.2$, $\alpha_r:0.1$, $(1 - \alpha_c - \alpha_f - \alpha_r):0.5$, so we can see that the sentiment distribution predicted by the text of target post has the highest weight 0.5, it makes sense, and other factors play the role for aiding the prediction.

| Liblinear L1R-LR | Basic | Context | Friend | Response | Combined |
|------------------|-------|---------|--------|----------|----------|
| Accuracy | 0.393 | 0.431 | 0.433 | 0.420 | 0.441 |
| F-score | 0.386 | 0.429 | 0.430 | 0.415 | 0.437 |

Table 4-7 The results after adding addition info, evaluate using emoticon-labeled testing data (D1), Basic means the base classifier.

We compare the improvement with other two approaches in the Table 4-8, given the D2 data and use accuracy to evaluate. The Feature engineering based approach uses the prediction emotion as features. We can see that the most improvement is adopting the Markov-transition based approach. Besides, In regard to the accuracy of basic sentiment classifier, the Liblinear L1R-LR is better than CRF model, so it's good if the basic sentiment classifier is replaceable so that we can replace current base classifier with better base classifier in the future.

| accuracy | Basic | context | friend | Response | Combined |
|---------------------------|-------|---------|--------|--------------|--------------|
| Feature engineering based | 0.549 | 0.555 | 0.553 | 0.551 | 0.556 |
| Graphical model based | 0.508 | 0.512 | 0.511 | 0.518 | 0.515 |
| Markov-transition | 0.549 | 0.558 | 0.569 | 0.560 | 0.570 |

| | | | | | |
|-------|--|--|--|--|--|
| based | | | | | |
|-------|--|--|--|--|--|

Table 4-8 the results of proposed three approaches in D2 data

We also try to apply this Markov-transition based model to binary classification for each emotion. Take anger as example, in training or testing data, all non-anger emotion are relabeled as non-anger, and train a binary classifier to predict the sentiment of a post belonging to anger or non-anger, and same for other sentiment. Note that the sentiment transition matrix becomes to learn the conditional probability of anger and non-anger, it may be not so meaningful than original six-class emotion, e.g. anger post may has responses containing anger, disgust, or surprise, but disgust and surprise are seen as the same non-anger emotion in here, so it may lose some detailed information. The experiment result shows in Table 4-9~ Table 4-14, the classifier and features are the same with Table 4-7, adopts Liblinear L1RLR and features mentioned in Section 3.1. The improvement for each sentiment in accuracy is in the range of 4.3%~0.2%, and the surprise has the most improvement.

(Detailed improvement in accuracy for each sentiment is that surprise(4.3%)

> disgust(1.9%) = joy(1.9%) > anger(0.7%) > fear(0.2%) = sadness(0.2%))

Besides, the performance (accuracy) of basic sentiment classifier for each sentiment is that joy(80.0%) > anger(77.4%) > sadness(70.0%)

> fear(69.0%) > disgust(63.0%) > surprise(61.8%), joy and anger are strong emotion which can easily be recognized, and surprise is not easily recognized because user may be surprise for positive or negative things, so it's hard to be recognized from the text. The disgust is similar emotion with anger, so it's not easily distinguished. Finally, the fear and sadness are also not easily judged because they are both triggered by unhappy things.

| Liblinear L1R-LR | Basic | Context | Friend | Response | Combined |
|---------------------|-------|---------|--------|----------|--------------|
| Accuracy | 0.774 | 0.780 | 0.778 | 0.779 | 0.781 |
| F-score | 0.774 | 0.782 | 0.779 | 0.781 | 0.784 |

Table 4-9 Binary classification for anger, the most improvement in accuracy is 0.7%,
and in F-score is 1%

| Liblinear L1R-LR | Basic | Context | Friend | Response | Combined |
|---------------------|-------|---------|--------|----------|--------------|
| Accuracy | 0.618 | 0.640 | 0.661 | 0.620 | 0.661 |
| F-score | 0.618 | 0.683 | 0.680 | 0.659 | 0.685 |

Table 4-10 Binary classification for surprise, the most improvement in accuracy is 4.3%,
and in F-score is 6.7%

| Liblinear L1R-LR | Basic | Context | Friend | Response | Combined |
|---------------------|-------|---------|--------|----------|--------------|
| Accuracy | 0.700 | 0.702 | 0.701 | 0.702 | 0.702 |
| F-score | 0.700 | 0.703 | 0.702 | 0.703 | 0.704 |

Table 4-11 Binary classification for sadness, the most improvement in accuracy is 0.2%,
and in F-score is 0.4%

| Liblinear L1R-LR | Basic | Context | Friend | Response | Combined |
|---------------------|-------|---------|--------|----------|--------------|
| Accuracy | 0.630 | 0.634 | 0.649 | 0.640 | 0.649 |
| F-score | 0.630 | 0.635 | 0.654 | 0.645 | 0.654 |

Table 4-12 Binary classification for disgust, the most improvement in accuracy is 1.9%,
and in F-score is 2.4%

| Liblinear L1R-LR | Basic | Context | Friend | Response | Combined |
|---------------------|-------|---------|--------|----------|--------------|
| Accuracy | 0.690 | 0.692 | 0.690 | 0.690 | 0.692 |
| F-score | 0.690 | 0.692 | 0.690 | 0.690 | 0.693 |

Table 4-13 Binary classification for fear, the most improvement in accuracy is 0.2%, and in F-score is 0.3%

| Liblinear L1R-LR | Basic | Context | Friend | Response | Combined |
|---------------------|-------|---------|--------|----------|--------------|
| Accuracy | 0.800 | 0.819 | 0.813 | 0.809 | 0.819 |
| F-score | 0.800 | 0.819 | 0.813 | 0.809 | 0.819 |

Table 4-14 Binary classification for joy, the most improvement in accuracy is 1.9%, and in F-score is 1.9%

Chapter 5. Conclusion

In this paper, the information from the three factors, i.e. temporal context, social, and response is verified helpful in sentiment detection. We propose three approaches for using the information. From experiment result, we know that after the method using the information of the three factors, the result can get a decent improvement than method without using, besides, the Markov-transition based approach is the most effective than other two approaches, and also contains the advantage of the other two approaches.

In the future, because that the accuracy of multi-class sentiment detection of the six emotion is not high enough, although it can achieve reasonable result competitive to state-of-the-art if transform the six emotion to positive/negative emotion, we still believe that exploring more detailed sentiment, e.g. anger, sadness, it will be helpful for sentiment analysis in the future. So we still can try to improve the result further for the problem. And about the context, friend factors, we may use topic detection to judge the

previous posts or friend's recent posts if they discuss the same topic with target post, it's helpful for bringing in more related posts. In addition, some of emotions are similar, for example, anger and disgust, we may try to model the sentiment detection as multi-label problem, i.e. the post can be labeled with multiple emotions.

Reference

- [1] Cheng-Te Li, Hung-Che Lai, Chien-Tung Ho, Chien-Lin Tseng, Shou-De Lin. 2010 Pusic: musicalize microblog messages for summarization and exploration WWW'10
- [2] Lun-Wei Ku and Hsin-Hsi Chen (2007). Mining Opinions from the Web: Beyond Relevance Retrieval. *Journal of American Society for Information Science and Technology*, Special Issue on Mining Web Resources for Enhancing Information Retrieval, 58(12), pages 1838-1850. Software available at <http://nlg18.csie.ntu.edu.tw/opinion/index.html>
- [3] FÅ Nielsen(2011), A new ANEW: Evaluation of a word list for sentiment analysis in microblogs, In International Workshop on Making Sense of Microposts 2011
- [4] Sun, Y. T.; Chen, C. L.; Liu, C. C.; Liu, C. L.; and Soo, V. W. 2010. Sentiment Classification of Short Chinese Sentences. In Proceedings of Conference on Computational Linguistics and Speech Processing (ROCLING'10), 184–198.
- [5] Mei-Yu Chen, Hsin-Ni Lin, Chang-An Shih, Yen-Ching Hsu, Pei-Yu Hsu, Shu-Kai Hsieh. 2010. Classifying Mood in Plurks. In Proceedings of Conference on Computational Linguistics and Speech Processing (ROCLING 2010), 172–183.
- [6] Go, A.; Bhayani, R.; and Huang, L. 2009. Twitter Sentiment Classification using Distant Supervision. *Technical Report*, Stanford University.
- [7] Ekman, P.: An Argument for Basic Emotions. *Cognition and Emotion*. 6, 169–200 (1992)

- [8] Barbosa, L., and Feng, J. 2010. Robust Sentiment Detection on Twitter from Biased and Noisy Data. In Proceedings of International Conference on Computational Linguistics (COLING'10), 36–44.
- [9] Bermingham, A., and Smeaton, A. F. 2010. Classifying Sentiment in Microblogs: is Brevity an Advantage? In Proceedings of ACM International Conference on Information and Knowledge Management (CIKM'10), 1183–1186.
- [10] Bifet, A., and Frank, E. 2010. Sentiment Knowledge Discovery in Twitter Streaming Data. In Proceedings of International Conference on Discovery Science (DS'10), 1–15.
- [11] Davidov, D.; Tsur, O.; and Rappoport, A. 2010. Enhanced Sentiment Learning Using Twitter Hashtags and Smileys. In Proceedings of International Conference on Computational Linguistics (COLING'10), 241–249.
- [12] J. Read. Using emoticons to reduce dependency in machine learning techniques for sentiment classification. In Proceedings of ACL-05, 43rd Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, 2005.
- [13] Long Jiang, Mo Yu, Ming Zhou, Xiaohua Liu and Tiejun Zhao. 2011 Target-dependent Twitter Sentiment Classification, ACL2011
- [14] P. Calais, A. Veloso, W. Meira Jr. and V. Almeida 2011. From Bias to Opinion: A Transfer-Learning Approach to Real-Time Sentiment Analysis. SIGKDD 2011.
- [15] Johan Bollen, Alberto Pepe, Huina Mao, 2010. Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. WWW2010
- [16] Stanford Log-linear Part-Of-Speech Tagger and Stanford Chinese Word Segmenter
- <http://nlp.stanford.edu/software/tagger.shtml>
- <http://nlp.stanford.edu/software/segmenter.shtml>

- [17] Changhua Yang, Kevin Hsin-Yih Lin, Hsin-Hsi Chen 2007. Emotion Classification Using Web Blog Corpora. WI '07 Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence
- [18] Strapparava, C., and Valitutti, A. 2004. Wordnet-affect: an Affective extension of wordnet. In Proceedings of International Conference on Language Resources and Evaluation, 1083–1086.
- [19] LingPipe
<http://alias-i.com/lingpipe/>
- [20] Weka
<http://www.cs.waikato.ac.nz/ml/weka/>
- [21] LibSVM and LibLinear
<http://www.csie.ntu.edu.tw/~cjlin/libsvm/>
<http://www.csie.ntu.edu.tw/~cjlin/liblinear/>
- [22] Pak, A., and Paroubek, P. 2010. Twitter as a Corpus for Sentiment Analysis and Opinion Mining. In Proceedings of International Conference on Language Resources and Evaluation (LREC' 10), 1320–1326.
- [23] Li, S.; Zheng, L.; Ren, X.; and Cheng, X. 2009. Emotion Mining Research on Micro-blog. In Proceedings of IEEE Symposium on Web Society, 71–75.